



# Tradeoffs in the Development of the SPASE Data Model

James R. Thieman<sup>1</sup>, Todd King<sup>2</sup> and D. Aaron Roberts<sup>3</sup>

<sup>1</sup>National Space Science Data Center, NASA/Goddard Space Flight Center

<sup>2</sup>Institute of Geophysics and Planetary Physics, UCLA

<sup>3</sup>Heliospheric Physics Laboratory, NASA/Goddard Space Flight Center



## ABSTRACT

Topic: Facilitating access to and retrieval of data across widely-distributed, disparate, science archives.

SPASE is the Space Physics Archive Search and Extract project. This project is intended to provide a common basis for locating and retrieving data of interest for the space and solar physics community across multiple widely-distributed heliophysics archives and data centers. Common terminology that maps to enable the disparate metadata being used by these data archives around the world enables unified searches and ready comparison of the results to determine time overlaps, data commonalities, applicability for research purposes, etc. The project has developed the SPASE Data Model for the description of heliophysics data sets. The success of this project depends on the wide usage of the Data Model in the community.

In this presentation we will talk about the development of the Model through dedicated international committee work and, in particular, the difficult tradeoffs that must be made. For example, to what level of detail should the model aim to describe data sets? Should just the finding and approach to acquisition of the data be supported through the model usage or should it describe data sets in sufficient detail that the user will obtain knowledge of all the parameters measured and details of resolution, cadence, etc. for the parameters? What tools will be needed to enable archives and data centers to map their existing metadata to the SPASE metadata in a reasonably automated manner? Version 1.2 of the Data Model is available presently and will soon be "frozen" for usage in a stable environment. (See <http://www.spase-group.org/>) The Model will evolve as the needs of the community dictate. We invite participation and feedback in the evolution as we make the decisions that affect the future of heliophysics science archiving.

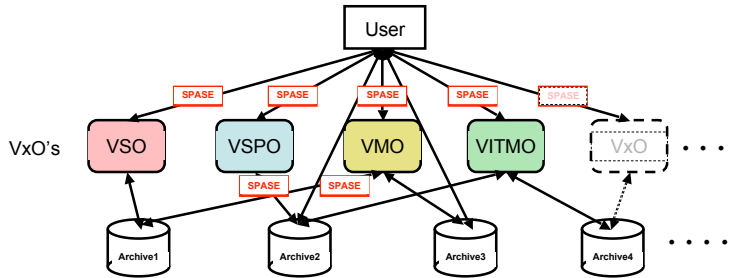
## What is SPASE?

### Space Physics Archive Search & Extract

An International, community-based standards organization with the goals of:

- Easing data search and retrieval across the Space and Solar Physics data environment
- Defining and maintaining a Data Model for Space and Solar Physics interoperability
- Demonstrating the Model's viability

## DATA ENVIRONMENT



Tradeoffs begin in the data environment.

- Users can access archives directly as they have before or they can use Virtual Observatories (VxO's) for access to multiple archives within their discipline
- SPASE can provide access across the disciplines in a unified way
- SPASE can be used for VxO communication with archives within the discipline or they may choose other protocols as needed.

## TRADEOFFS

### Finding Data

Description only for "Full Dataset" Level  
Automated Metadata Mapping  
Cross-Discipline Keywording  
Conceptual Metadata  
Describing Objects

### Using Data

Description to "Granule" Level  
Manual Metadata Mapping  
Intra-Discipline Keywording  
Structural Metadata  
Describing Bytes

### Finding Data vs. Using Data

- Should the SPASE Data Model support just finding data or using data as well?
- If using data is desired, is this just for data documentation or for data services?
- Will "finding" vs "using" be different in different disciplines?

### Automated vs. Manual Metadata Mapping

- Automated metadata mapping from native descriptions to SPASE descriptions is quickly complicated by increasing level of detail
- What number of datasets to be mapped make it worthwhile developing an automated mapping tool?
- Tools developed for mapping metadata so far indicate a mixed automated/manual approach is best

### Conceptual vs. Structural Approaches

- "Conceptual" metadata is usually defined by variable physical conditions or generic terms such as the location of a "bow shock"
- "Structural" metadata is usually defined in a "fixed" or "specific" way such as a geometrical definition of a boundary
- Conceptual metadata can result in confusing overlapping of keywords, especially across disciplines, but usually is better for a specific discipline

## Summary

- These tradeoffs profoundly influence the evolution of the data model and the SPASE effort
- The community, in describing datasets, and using SPASE-based searches provides the feedback for choices among the tradeoffs
- We appreciate any level of commitment to this effort that you care to make

## SPASE Data Model



<http://www.spase-group.org>

